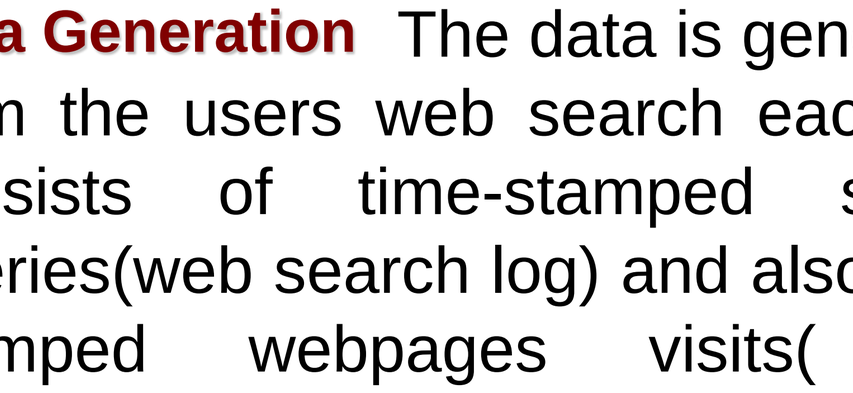
Web Search and Browser Log analysis using BangDB for Decision Support

Ravi teja Ravinoothala,
Department of Computer Science
University of Bridgeport, Bridgeport, CT

Abstract Bigdata is an emerging data that is beyond the capacity of a conventional database tools to analyse.BangDB rides the bigdata where data is processed using clusters. Web server logs are semi-structured files generated in a large volume by web users usually of flat text files. This paper performs the analysis on log files using BandDB in a distributed cluster. This framework affective detects the useful statistics like top n searches in a time period by the unique users search and pages accessed by the users. These results are compared with other NoSQL databases, and it results in a better time efficiency, storage and processing speed.

Data Generation The data is generated from the users web search each row consists of time-stamped search queries(web search log) and also time-stamped webpages visits(web browser log).Initial data is filtered with queries that contain the phase "buy online" at least 2 times, this filter is employed to make the study relevant to online shopping.



Conclusion

The BangDB gives the highest IOPS (Input/output Operations Per Second,) in several scenarios when compared to Oracle's BerkleyDB and Google's LevelDB.

Design and Implementation

Sliding window: In real time analysis, we are interested in most recent data and wish to analyze the data accordingly. BangDB provides sliding table concept, which means that we can simply create a table providing time range and then work within the time range as the window always keeps on sliding continuously. Here we strictly want to work within the defined recent window.

Counting In almost all analytical purposes, counting in inevitable. Again these counting could be counting since beginning or for specified time window which keeps sliding. For such use cases, BangDB provides native constructs for counting.

TopK TopK means keeping track of top k items. This is another important feature from analytics perspective. TopK has been a topic of interest for many researchers and analysts and therefore used at many places. BangDB provides native construct for TopK. TopK can again be done in absolute manner or within a sliding window. These are available in BangDB as fully baked up constructs and hence be used directly. However we can enable different analytical capabilities using BangDB different features.